Practical *versus* pragmatic: enlarging the selection task, extending reasoning

Romain Boissonnade (romain.boissonnade@unine.ch)
University of Neuchâtel, Institute of Psychology and Education, Espace Agassiz 1,
2000 Neuchatel, Switzerland

Antonio Iannaccone (antonio.iannaccone@unine.ch)
University of Neuchâtel, Institute of Psychology and Education, Espace Agassiz 1,
2000 Neuchatel, Switzerland

Abstract

In the selection task, individuals generally do not follow the deductive rules of standard logic. In the present research, a new abstract selection task was designed by extending the range of cards that the participants face (inspired by Manktelow, Sutherland and Over, 1995). It was used in two experiments to test predictions about the matching bias and the probabilistic approach of reasoning. By multiplying the number of cards, we showed a reduction of the *pq* response, indicating that the matching bias may partly be due to practical features of the task (experiment 1). Surprisingly, half of the participants unsystematically turned over specific cards in some categories. The post hoc justifications allowed us to distinguish several possible interpretations of the task, and differing strategies, (either deductive or inductive) in a bid to uncover contradictions. The result was replicated with a short-time procedure. It also showed a progression of the logical *p not-q* response (experiment 2). We thus propose that a distinction be made between a pragmatic and a practical component of the matching bias. The features of the task also define a range of deductive and inductive strategies to solve the problem.

Keywords: reasoning; conditional; logic; matching bias

Introduction

Several theories compete to explain the peculiarities of human thinking and reasoning facing logical tasks. The Wason's selection task (1966, 1968) is considered to be a paradigmatic deductive task aimed at studying these cognitive processes and is still widely used today. In this task, the participant is informed about a game where a card always has one letter on one face and a number on the other. A conditional rule is stated: *"If there is a vowel on one side of the card, then there is an odd number on the other side"*. Four cards are proposed: a vowel (antecedent *p*), a consonant (*not-p*), an odd number (consequent *q*) and an even number (*not-q*). The task is to select the cards which determine whether the rule is true or false. Following standard logic, there is only one correct response: turning over the *p* and *not-q* cards. It has been consistently shown that the selection task elicits correct responses only 5-10% of the time, and that most of the subjects choose to turn over only the *p* or the *p* and *q* cards, thus showing what appears to be a confirmation bias (Wason, 1966, 1968). Although there are various theoretical perspectives, one question that is common to all of them is to ask where the unexpected responses originate. In this exploratory paper, a variation of the task is developed in order to investigate the role of the external features of the task.

Pragmatic ecology of the selection task

One of the most prevalent criticisms of the selection task is about its lack of ecology. Several researches thus involved experimental variations on the content of the cards or in the instructions. It had been argued that responses are more likely to conform to standard logic if a thematic content is provided. However Cheng and Holyoak (1985) explained this facilitation effect by a pragmatic stance: most of the thematic versions of this task imply deontic reasoning, i.e. people have to reason about possibilities rather than truth validity, looking for items that violate the rule rather than checking the truth of the rule. Their investigation opened an avenue for other pragmatic analyses, mainly based on the idea that the task is thought to depend on the specific points of view and goals of the participants themselves. The "relevance effect" is another interesting pragmatic influence. Sperber, Cara and Girotto (1995) highlighted the importance of the communicational context of the problem: tacit rules of conversation are used by the participant to interpret the task given by an experimenter. Evans (2014) also explained pragmatic cues as being responsible for the common *pq* response. Indeed, Evans and Lynch (1973) showed that including a negation in the conditional rule does not change the response patterns: people still concentrate on the aforementioned *p* and *q* items, neglecting the negation. Thus Evans regularly defined the "matching bias" as a tendency of people to focus on the items mentioned in the conditional rule. This matching bias supposes basic cognitive processes prior to reasoning.

The shape of the task: toward a practical ecology

Evans (2014) developed the "dual process" theory of reasoning. A first heuristic system, System 1, is largely automatic, fast and unconscious. It is largely pragmatically focussed and pre-empts the analytic system, System 2, which is slower and more conscious. Influenced by distributed and extended cognition approaches, we hypothesize that other contextual factors may colour cognitive processes, in particular whether "practical" cues, (i.e. potential actions supported by the objective features of the task), may impede System 1 functioning, rather than "pragmatic" cues (i.e. the linguistic aspects like the

instructions, or the contents of the cards). For instance, the classical selection task, mobilizes four instances, based on a Boolean categorisation of the logical system, i.e. truth tables, and one single item is provided for each instance. It is difficult to find problems in everyday life that imply having to choose from one *p*, one *not-p*, one *q* and one *not-q* items. Hence, one way to question the role of the shape of the task is to consider the number and diversity of cards that participants have to deal with. In other words, the classical "shape" of the task (i.e. four cards) could limit the System 2 reasoning processes of the agent who has to think about "pre-digested" or "pre-categorized" information.

Large array selection tasks

Manktelow, Sutherland & Over (1995) proposed a thematic Large Array Selection Task (LAST) with an immigration regulation theme: participants had to pretend to be an airport inspector, checking boarding passes to see whether passengers were transiting or entering a territory (*p* and *not-p*), and whether they were vaccinated against cholera or not (*q* and *not-q*). The range of cards was extended to include a third piece of information - the nationality of the passengers. The authors also provided claims about geographical origins and risks. In four experiments, they showed that subjective probabilities and expected utilities triggered a *perspective effect* based on the participant's point of view of immigration and risks. It appeared that knowledge about the probability of the events played a crucial role in deontic reasoning. Similarly, Fairley, Manktelow and Over (1999) used a thematic LAST and proposed a comparable explanation in a causal context, where knowledge about the probability of finding alternative causes and disabling conditions drastically changed reasoning on factual laws. It appears then, that the context complexity is predetermined by subjective a priori knowledge about this context.

To better ascertain the role of the array of cards in the cognitive processes, a thematic LAST should be replaced by an abstract one. As far as we know, no such study exists. However some studies, related to decision-making and expected utilities, suggested the importance of strategies based on Bayesian rationality in order to reduce uncertainty. Oaksford, Chater, Grainger & Larkin (1997) and Oaksford, Chater & Grainger (1999) used abstract selection tasks with sequential sampling: participants were repeatedly presented with cards extracted from a larger deck. In this case, people chose specific cards to gain information about the rule, noticeably by turning over the *p* and *q* cards because of their supposed poor frequency into the deck (Oaksford & Chater, 2003). The repeated resolution is crucial here in order to observe probabilistic strategies. However, if the repeated nature of the task is removed, it is rather difficult to know how multiplying the cards in a selection task may orient choices. Following the probabilistic approach, the response frequencies should be comparable with the classical task if the proportions of *p*, *not-p*, *q* and *not-q* cards are identical. An abstract LAST could then be an interesting test for the probabilistic perspective.

Understanding the strategies

In order to study the effects of this enlarged abstract selection task and to have a fine-tuned interpretation of the response patterns, we followed Elqayam and Evans' recommendations (2011). A mixed methodology was used to investigate how individuals are confronted to an enlarged selection task. We first asked them to make a selection and then to justify their choice. This enables us to understand the responses and the strategies people may possibly use. This analysis is particularly driven by the assumption that the participants can use an inductive reasoning. Indeed, following the probabilistic approach, people try to become more confident and « *inductive strength is probably more important in determining people's behaviour than deductive correctness, even on putative deductive reasoning tasks* » (Oaksford & Hahn, 2007, p. 271). The creator of the selection task was himself aware of this: *"The subjects […] may, in fact, have regarded the cards as items in a sample from a larger universe, and reasoned about them inductively rather than deductively"* (Wason & Shapiro, 1971, p. 70). Thus the task could be performed with inductive strategies, possibly related to the "shape" of the task. In the following study, the participants had to solve an enlarged selection task and then provide explanations.

Study 1

We hypothesized that using an enlarged selection task would change the types of response (H1). Following most research in the domain, we wonder whether the frequency of the *p not-q* responses may increase in comparison to Wason (1968) (H2). In order to explore these effects, the participants had finally to justify their response.

Method

Participants Bachelor and master students of the faculty of Humanities and Social Sciences in Neuchâtel (Switzerland) participated voluntarily (*N* = 46; *Mean age* = 23;0; *male/female* = 17/29).

Material We used an enlarged selection task including 16 cards, with 4 cards per instance (*p, not-p, q, not-q*), and in each of the instances, one card was repeated (e.g. two "A" cards among the vowels). The arrangement of the cards was counterbalanced in four different orders. The instructions were provided as follow, translated into French: *"Exercise (about 5 minutes): Here is a card game. All the cards have one letter on one face and a number on the other face. We place a few cards of this game before you. One rule states that if there is a vowel on one side of a card, then there is an odd number on the other side of the card. Among the cards in front of you, which card(s) need to be turned over in order to know whether the rule is true or false?"* (see Appendix). The time period (5 minutes) is provided to ascertain that the participants take enough time to look at the cards. On the back of the page, the participant writes a justification (*"Could you justify your response? Why did*

you choose this(these) card(s)?"), and then explains his/her method (*"Could you describe how you reasoned?"*). Once the participant begins to give their justification, they can no longer change their card choice. Finally, we asked the participant if they had ever seen a similar kind of task with cards. The results of those who recognized the task were withdrawn before analysis began.

Procedure The participants had free time to read and solve the task, and then to turn the page over and justify.

Results

Response analyses There is no statistical difference in the frequency of the logical *p not-q* response (*ns*) (H1 rejected). Nevertheless, there is a significant difference in the distribution of the three main classes of responses (*p*; *pq*; other), $\chi^2(2, N = 46) = 18.203$, $p < .001$. Follow-up analyses with residuals indicate that: more *p* response are provided (37% vs 21%, $z = 7,3$) as well as "other" responses (41% vs 26%, $z = 7,0$); *pq* responses diminished importantly (22% vs 53%, $z = -14,4$) (H2 corroborated) (see table 1, fig. 1).

Table 1: Selections in exp. 1 and in Wason's study (1968)

Study	Selection				N
	P	PQ	P-Q	Other	
Current study	17 (37%)	10 (22%)	3 (6%)	16 (35%)	46
Wason (1968)	7 (21%)	18 (53%)	1 (3%)	8 (23%)	34

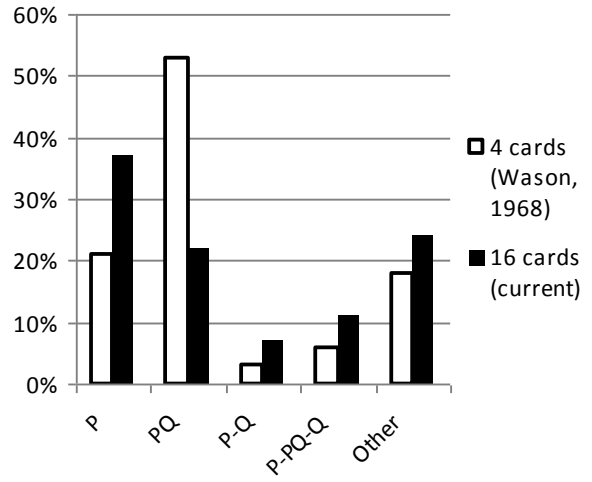

Figure 1: Proportions of selections in Wason's study (1968) and in exp. 1

Less participants chose the *pq* response. This suggests that the matching bias is less effective, and the objective shape of the task partially accounts for this bias. We propose to distinguish a "pragmatic matching bias", due to the stated rule, as claimed by Evans (2014), and a "practical matching bias", due to the presence of a number of cards. In case of four cards, the shape of the task invites people to actively pair the cards. In our task, this "practical matching bias" is removed. The *pq* responses are due to a residual pragmatic bias (the rule still focuses on *p* and *q* cases).

Unexpected response diversity Surprisingly, a new class of response appeared in the 16-cards condition, involving greater response heterogeneity than in Wason's study. Two classes of responses can now be distinguished, depending on whether the participants selected the cards "systematically" or "unsystematically". For each instance, they were expected to select all or none of the cards in a category. Indeed, some people made purely "systematic" selections (i.e. only selecting 4 cards in one or several of the instances) ($N = 24$). But 22 participants (48%) made "unsystematic" selections (i.e. choosing only one, two or three cards among four).

1 – Unsystematic selections are common in any card combination. In particular, 4 of the 17 participants gave an unsystematic *p* response, 2 participants of 9 gave an unsystematic *pq* response, and 1 over 3 gave an unsystematic *p not-q* response.

2 – In this "unsystematic" category, some "partial" selections are made (few cards selected in specific instances) ($N = 18$). Some people also made "purely partial" selections by choosing few cards in each of the four instances ($N = 4$). No-one gave a mix of partial and systematic selection (few cards in some instance, all cards selected in other instances) ($N = 0$).

3 – All instances (*p*, *not-p*, *q*, *not-q*) are confronted to both selection styles, systematic or not. The relative importance of unsystematic over systematic responses is particularly increased for *not-p* and *not-q* cards: participants generally selected these cards unsystematically (see fig. 2).

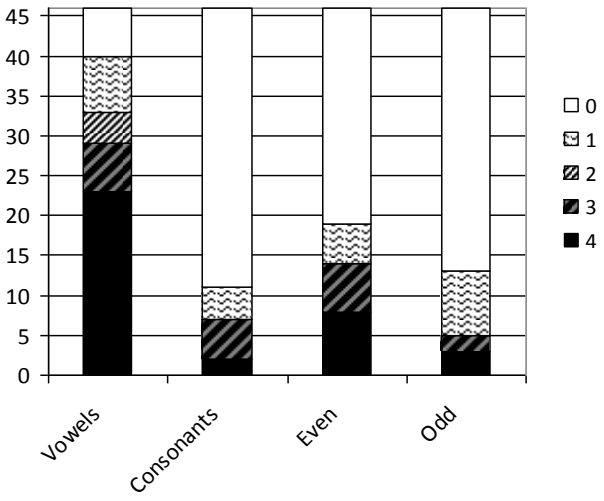

Figure 2: Number of participants per instance and per number of cards selected in exp. 1

Kinds of partial response Several unsystematic strategies can be distinguished: (i) choosing only one card in each

desired instance (ex : “A”, “K”, “3”); (ii) choosing only different cards in the desired instances (ex : “A”, “I”, “U”, “K”, “D”, “P”); (iii) choosing only repeated cards in the desired instances (ex: two “A” and two “2” cards). Finally, 4 participants (over 22) mixed up these strategies.

Post hoc justifications Written justifications are a first means to understanding the reasoning behind these strategies. The participants giving an unsystematic *pq* response (N = 3) justified their strategy in reference to the repeated cards. For instance: *“We need to look at all the different cards (the other A or 2 are not chosen) to confirm the rule that some cards can follow the rule and others not”*.

Among the unsystematic *p* responses (N = 4), one participant selected only two “A” cards and explained that he first considered the rule as true, then chose one “A” card among the vowels, and finally another “A” to check again. Another participant only chose three different vowels and did not explain the exclusion of the repeated “A” card. Another chose three different cards explaining that she expected specific even numbers related to the letters, seemingly looking for a meaning to this abstract task; she thus tried to discover a different underlying rule rather than directly dismissing the one stated in the instructions.

One participant chose only one of each repeated “A” and “3” cards, giving an unsystematic *p not-q* response (N = 1): *“there is no need to check all the cards with a vowel for instance, as we know that every time there is one vowel, then, on the back there is an even number”*. This ambiguous explanation suggests that he tried to know whether all the cards are following this rule or following another one.

Remaining questions In this experiment, people tended to use undocumented strategies, apparently illogical but reasonable. Rather than just checking the validity of the stated rule, they may use strategies to discover whether there is another rule governing the cards that possibly contradicts the stated rule, or any other specific meaningful rule instead. It is perhaps possible that the time allocated in the first study and the complexity of the task could explain the surprising response patterns.

Study 2

For a better comparison, we balanced two groups, some participants having the classical 4-card task and others having the 16-card task. We hypothesized a difference in the response distributions (H1). The 16-cards task would imply more *p not-q* response (H2). Following the first study, we hypothesized a diminished *pq* frequency in the 16-cards (H3). Finally, if unsystematic choices are due to a lack of time to observe the cards and give a response, shortening the time constraint would imply more “partial” responses (H4).

Method

Participants Ninety-eight french speaking students of a swiss Higher Teacher Training Institute participated voluntarily (N = 98; *Mean age* = 28;6 ; *male/female* = 34/64): 50 participants were randomly involved in the 4-cards condition, and 48 participated in the 16-cards condition.

Material The same material was used as in the first study.

Procedure The participants had one minute only to individually read and solve the problem. The experimental conditions occurred at the same time in eight separated rooms. Both experimental conditions occurred in each room.

Results

Response analyses Grouping the responses in three main classes (*p*, *pq*, other), no significant difference was found in the distributions of the conditions, χ^2 (2, N = 98) = 5,032, p = .081 (H1 not corroborated, see table 2, fig. 3)[1].

Table 2: Selections by experimental conditions in exp. 2

Study	Selection				N
	P	PQ	P-Q	Other	
4 cards	8 (16%)	24 (48%)	1 (2%)	17 (34%)	50
16 cards	8 (17%)	13 (27%)	7 (15%)	20 (41%)	48

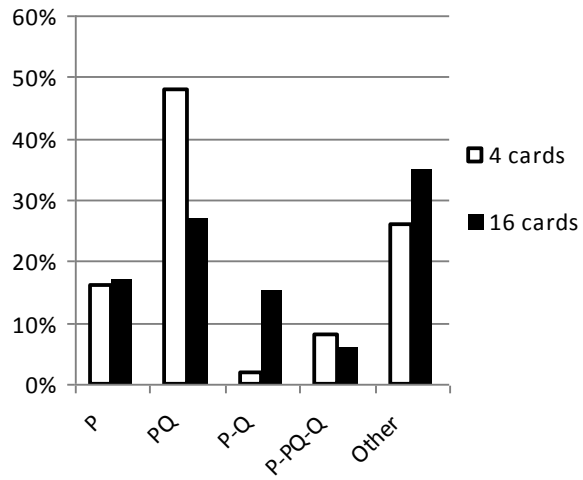

Figure 3: Proportions of selections in exp. 2

Interestingly, there are more *p not-q* selections in the 16-cards (N = 7) than in the 4-cards group (N = 1), p = .026 (one-sided Fisher’s exact test, *Cramer’s V* = 0.230) (see fig. 3). With more subjects, a progression of the *p not-q* expected response appeared (H2 corroborated). The observed strategies for this *p not-q* response are similar to those reported in Study 1 for the partial responses: (i) five participants unsystematically circled a few *not-q* cards,

[1]The distribution of the 4-cards condition does not differ with Wason’s (1968) distribution, *ns*. Identically, the responses of the 16-cards conditions do not differ from study 1, *ns*.

suggesting that they did not apply a purely deductive reasoning or that they did so with uncertainty; two of them circled a single *p* and a single *not-q* card; (ii) one participant chose two different *p* cards, and three different *not-q* cards, thus avoiding selecting the repeated cards; similarly, another selected one of the repeated *p*, and two different *not-q* cards (one of the repeated cards); (iii) one participant selected the repeated *p* and *not-q* cards.

As in study 1, there was a reduction of *pq* selections, $p = .027$ (one-sided Fisher's exact test, *Cramer's V* = 0.216). This pattern is comparable to study 1, with a decrease from 50 to 25% (H3 corroborated). This result apparently confirms a matching bias reduction and the time allocation does not particularly change this tendency. Further experiments may reveal small differences.

Description of unsystematic strategies Among the 16-card group, one can find systematic selections ($N = 22$), and unsystematic selections ($N = 26$). The proportions are not significantly differing from study 1, *ns* (H4 not corroborated). The proportions of unsystematic selections (54%) are very similar to Study 1 (48%). This makes it difficult to account for these unsystematic strategies by time pressure. As a consequence, it has to be supposed that the enlarged task reveals differentiated problem interpretations. In the partial responses, one can distinguish either purely unsystematic selections (selecting few cards in certain instances) ($N = 21$) or a mix of unsystematic and systematic selections (few cards selected in some instance, and all cards in other) ($N = 5$). Further investigations are currently under process.

Discussion

This study set out to show that the performances to the Wason selection task are partly due to its objective features. In line with a half century of literature, very few participants gave the expected *p not–q* response (less than 15%), either systematically (selecting all *p* and *not–q* cards) or unsystematically (selecting only a few *p* and/or *not–q* cards). Extending the objective part of the selection task does not radically imply more logical responses, although a small effect did arise. The 16-card selection task involved a little progression of this expected response, and this may reinforce the idea of a non-ecological 4-card selection task. When presented with more items, a few people tended to choose *p* and *not-q* cards, often unsystematically, potentially using inductive reasoning. Nevertheless, analysing the other responses is informative on the role of the task's shape.

Pragmatic and practical matching biases

Generally explained as a matching bias effect or as bi-conditional understanding, we observe a reduction of *pq* responses when there were more numerous items at the participants' disposal. In our opinion, this means that the problem was understood differently, not only due to the items mentioned in the conditional statement, but also because of the objective context the statement refers to. This allows a supposed two-sided matching bias. Indeed the situation for the subject is shaped by attentional cues, which here is the rule, but equally by the contextual structure. We therefore propose to distinguish the two dimensions, namely a "pragmatic component" and a "practical component" of the matching bias. This practical component, which is not located in the individual, also explains why a "matching bias" is often considered as robust in spite of instructional variations (e.g. Stahl, Klauer & Erfelder, 2008). Further investigations are needed to better understand how these components combine into an extended cognitive system, where individual cognition is in interaction with other elements like props.

Two rationalities, multiple strategies

We noticed an increased heterogeneity of the responses, larger than in the traditional selection task. Among them, some surprising responses occurred, which do not enter *a priori* in any category of logical argument: about 50% of the participants decided to turn over specific cards in some instances. These "partial" responses are diversely justified. Rather than being a random procedure due to external constraints and heuristic processes, these responses appear regularly, with or without time pressure, as a consequence of several unsystematic but reasonable strategies. Two main reasons support the use of these unsystematic strategies. Firstly, partial selections can be intended to explore the problem with a view to "seeing later", i.e. find out something else about the context and the cards, about the hidden meanings of the rule, about the experimenter's intentions… Secondly, these selections could be the result of a specific interpretation of the problem and of resulting deductive reasoning. Some participants supposed that all the cards of the game are governed by a rule which is potentially different from the rule stated by the experimenter: the participant has thus to know whether all the cards follow the stated rule or if they follow a different hypothetical rule, which may contradict the stated rule. In this interpretation, it would not be relevant to turn over all the cards of a category, as they are supposed to follow a common rule. Turning over a single vowel or a few vowels could be sufficient enough to determine the existence of an unknown rule of the game. We thus observed unsystematic strategies driven either by inductive or by hypothetico-deductive rationalities. We did not find a way to explain our results using a probabilistic approach, though we agree with Oaksford and Hahn (2007) about the presence of inductive processes in the selection task resolution. Further experiments should confirm these strategies and separate a potential rationalization effect.

What we learned about the selection task

It is possible that these deductive as well as inductive rationalities and these strategies were also used with the classical selection task. Indeed, they would be confounded, due to the limited range of possible responses. In Wason's selection task, only one card is proposed for each logical

instance and is considered as representative (for a logician). Hence, when a participant selects a card, we cannot know if he wants to select a category rather than a single case. It is supposed that the analysis confounded differing rational interpretations and reasoning strategies on the basis of their common final response. Doing so, we lack a fine understanding on how cognition proceeds. In short, processes are hidden by one-off responses.

Following Boissonnade, Tartas & Guidetti (2014), we suggest: 1) to complexify the selection task (here with the multitude of cards), as ecological situations are usually richer than logical pre-categorisations; 2) to go beyond performances and closely analyse reasoning and behaviour using a pragmatic stance and renewed techniques; 3) to refer to a sociocultural perspective in order to better understand meaningful interpretations that orient reasoning. New studies are currently in progress, using this enlarged task to track attention and reasoning with eye-tracking techniques and mixed methodologies.

Acknowledgments

We are especially grateful to Daniel Dukes (Center of Cognitive Science at University of Neuchâtel) whose generous suggestions and support have invaluably contributed to the article.

References

Boissonnade, R., Tartas, V., & Guidetti, M. (2014). Toward a cultural-historical perspective on the selection Task. *Integrative Psychological and Behavioral Science*, *48*(3), 341-364.

Cheng, P. & Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences, 34*(5), 233–248.

Evans, J. St B. T. (2014). Reasoning, Rationality and Dual Processes: Selected Works. London: Psychology Press.

Evans, J. St B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*(3), 391–397

Fairley, N., Manktelow, K. I., & Over, D. E. (1999). Necessity, sufficiency, and perspective effects in causal conditional reasoning. *Quarterly Journal of Experimental Psychology, 52A*, 771–790.

Manktelow, K. I., Sutherland, E. J., & Over, D. E. (1995). Probabilistic factors in deontic reasoning. *Thinking and Reasoning, 1*(3), 201–220.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5*(3), 193–243.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal Data Selection in the Reduced Array Selection Task (RAST). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(2), 441-458.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review, 10*, 289–318.

Oaksford, M., & Hahn, U. (2007). Induction, deduction and argument strength in human reasoning and argumentation. In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge, MA: Cambridge University Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 52*, 3-39.

Stahl, C., Klauer, K. C. & Erdfelder, E. (2008). Matching bias in the selection task is not eliminated by explicit negations. *Thinking & Reasoning, 14*, 281–303.

Wason, P. C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 273-281.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*, 63–71.

Appendix

Exercice (environ 5 minutes):

On a un jeu de cartes. Toutes les cartes comportent une lettre d'un côté et un chiffre de l'autre. On pose quatre cartes. Une règle indique : « s'il y a une voyelle sur un côté d'une carte, alors il y a un chiffre pair sur l'autre côté de la carte ». Parmi les cartes posées, quelle(s) carte(s) faut-il absolument retourner pour savoir si la règle est vraie ou fausse) (Entourez-la/les).